

Refer #7

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 03-177899

(43)Date of publication of application : 01.08.1991

(51)Int.Cl.

G10L 3/00

(21)Application number : 01-316991

(71)Applicant : FUJITSU LTD

(22)Date of filing : 06.12.1989

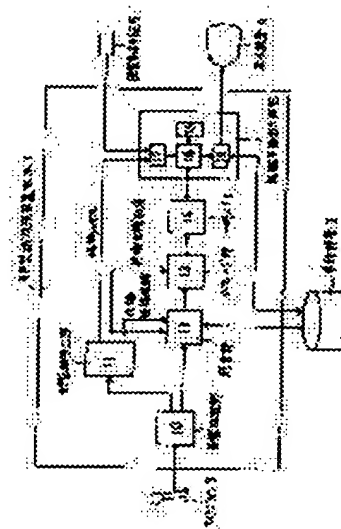
(72)Inventor : KIMURA AKIHIRO

(54) SPEECH RECOGNITION PROCESSOR

(57)Abstract:

PURPOSE: To recognize an extremely large vocabulary in real time by calculating the similarity of a word to be recognized according to use frequency and outputting a recognition result according to the similarity which is found so far at the point of time of the detection of a voice section and at a requested point of time.

CONSTITUTION: An output part 15 specifies and outputs words with high similarity according to a sorting result obtained up to the point of time of the end determination of the voice section at the point of time of the end determination. Further, words which have high similarity corresponding to the current sorting result except words which are already outputted are specified and outputted. Consequently, many words can be recognized in real time without making the operator have a feeling of physical disorder.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

⑨ 日本国特許庁(JP)

⑩ 特許出願公開

⑫ 公開特許公報(A)

平3-177899

⑤ Int. Cl.⁵

識別記号

庁内整理番号

⑬ 公開 平成3年(1991)8月1日

G 10 L 3/00

3 0 1 D

8842-5D

審査請求 未請求 請求項の数 3 (全13頁)

⑭ 発明の名称 音声認識処理装置

⑮ 特 願 平1-316991

⑯ 出 願 平1(1989)12月6日

⑰ 発 明 者 木 村 晋 太 神奈川県川崎市中原区上小田中1015番地 富士通株式会社
内

⑱ 出 願 人 富 士 通 株 式 会 社 神奈川県川崎市中原区上小田中1015番地

⑲ 代 理 人 弁 理 士 森 田 寛 外2名

明 細 書

1. 発明の名称

音声認識処理装置

2. 特許請求の範囲

(1) 電気信号に変換された音声信号の特徴時系列データを算出する音響処理部(10)と、音声信号のもつパワー値に従って音声区間を検出する音声区間検出部(11)と、認識対象の単語のもつ特徴時系列データを管理する単語辞書(2)と、検出された音声区間により抽出される上記特徴時系列データと上記単語辞書(2)の特徴時系列データとを照合することで、入力された音声信号と認識対象の単語との間の類似性を評価する照合部(12)と、評価された類似性に従い類似性の高い順に従って認識対象の単語をソーティングするソーティング部(13)と、ソーティングされた単語の内の上位の高い類似性をもつ1つ又は複数の単語を表示出力する認識単語出力部(15)と、該認識単語出力部(15)

に対して表示出力処理を要求する検索要求手段(5)とを備える音声認識処理装置において、

上記認識単語出力部(15)は、音声区間の終端確定時点において、その時点までに求められているソーティング結果に応じて上位の高い類似性をもつ1つ又は複数の単語を特定して出力するとともに、上記検索要求手段(5)からの検索指示があるときに、その時点までに求められているソーティング結果に応じてそれまでに出力した単語を除いた上位の高い類似性をもつ1つ又は複数の単語を特定して出力するよう処理することを、

特徴とする音声認識処理装置。

(2) 請求項(1)記載の音声認識処理装置において、

単語辞書(2)は、使用頻度の大きい順に従って単語を管理するとともに、照合部(12)は、この単語の並び順に従って類似性を評価するよう構成されてなることを、

特徴とする音声認識処理装置。

(3) 請求項(1)、(2)記載の音声認識処理装置において、

次の音声信号の入力若しくは入力要求があるときには、照合部(12)は、実行中の類似性の評価処理を中止するとともに、ソーティング部(13)は、実行中のソーティング処理を中止して求められているソーティング結果をクリアするよう構成されることを、

特徴とする音声認識処理装置。

3. 発明の詳細な説明

(概要)

音声信号の表す単語を実時間で認識する音声認識処理装置に関し、

非常に多くの単語をオペレータに違和感を与えることなく実時間で認識することを目的とし、

音声信号の特徴時系列を算出する処理部と、音声区間を検出する検出部と、認識対象の単語の特徴時系列を管理する単語辞書と、音声区間の特徴時系列と単語辞書の特徴時系列とを照合して類似性を評価する照合部と、認識対象の単語を評価された類似性に従ってソーティングするソーティン

グ部と、上位の高い類似性をもつ単語を表示出力する出力部と、出力部に対して出力要求する要求手段とを備える音声認識処理装置において、出力部は、音声区間の終端確定時点において、その時点までのソーティング結果に応じた上位の高い類似性をもつ単語を特定して出力するとともに、要求手段からの指示があるときに、その時点のソーティング結果に応じた上位の高い類似性をもつ単語であって既出力のものを除く単語を特定して出力するよう構成する。

(産業上の利用分野)

本発明は、入力される音声信号の表す単語情報を実時間で認識する音声認識処理装置に関し、特に、非常に多くの単語情報をオペレータに違和感を与えることなく実時間で認識できるようにする音声認識処理装置に関するものである。

近年、音声による文書の作成処理や音声による商品名の入力処理等といったように、産業界の様々な分野で発声された音声の表す単語を実時間で

認識して出力していく音声認識処理装置が用いられるようになってきている。このような音声認識処理装置では、より多くの単語を認識対象とする要求が強まりつつあるので、この要求に対応できるための手段を講じていく必要がある。

(従来技術)

音声認識処理装置では、認識対象となる単語の特徴時系列データ(スペクトルの時系列データ等)を管理する単語辞書を用意するとともに、処理対象の音声信号の音声区間を検出して該音声区間の音声信号の特徴時系列データを検出するよう構成し、この検出された音声信号の特徴時系列データと単語辞書に管理される特徴時系列データとの間の類似性を類似度や確率や距離等に従って評価して、上位の高い類似性をもつ1つ又は複数の単語情報を認識結果として出力するという構成を採ることになる。ここで、類似度や確率で類似性を評価するときにはその値が大きい程類似性が高くなり、距離で類似性を評価するときにはその値が小

さい程類似性が高くなることになる。

ここで実行される音声区間の検出処理は、具体的には、音声信号のパワーがある判定レベル値以上に入ることを検出することで音声区間の始端を検出するとともに、ある判定レベル値以下に下がることを検出することで音声区間の終端候補を検出する。そして、音声信号のパワーのレベルがこの終端候補に続いて所定時間(通常、0.3秒程度に設定される)の間判定レベル値以下にあることを検出することで音声区間の終端を検出するとともに、この所定時間内に音声信号のパワーのレベルが再び判定レベル値に上昇してしまう場合には、それまでの音声区間が継続しているものと判断して検出した終端候補を取り消していくことで、音声区間の終端の検出を実行することになる。

第7図に、この音声区間の検出処理機能の状態遷移図を図示する。図中、“S0”が無音声区間の状態、“S1”が有音声区間の状態、“S2”が終端判定区間の状態を示している。音声区間の検出処理は状態“S0”からスタートする。この

状態“S0”では音声パワーが閾値 P_{th} を越えるかどうかを監視しており、越えると状態“S1”に遷移する。このときの遷移時刻 t_1 が音声区間の始端の時刻であり、 s_t というレジスタに格納される。状態“S1”では、音声パワーが閾値 P_{th} を下回るかどうかを監視しており、下回ると状態“S2”に遷移する。このときの遷移時刻 t_2 が終端候補の時刻であり、 e_{tmp} というレジスタに格納される。状態“S2”では閾値時間 L_{th} （これが上述の0.3秒である）内に音声パワーが閾値 P_{th} を越えるかどうかを監視することで、無声破裂音や破裂音の前の促音が単語中にあることに基づき一時的な音声パワーの低下であるのかどうかを監視しており、音声パワーが閾値時間 L_{th} 内に閾値 P_{th} を越えるときには一時的な音声パワーの低下であったと判断して状態“S1”に戻るべく遷移する。このとき、レジスタ e_{tmp} がクリアされることで終端候補が取り消される。そして、更に、状態“S2”ではこの状態が閾値時間 L_{th} 継続するかどうかを監視することで、検出された

て表示してオペレータに通知していくとともに、オペレータがこの表示中に発声した単語に対応する単語情報を見出せないときには、用意される候補選択スイッチを押させるよう指示して、この指示に応じてオペレータがその候補選択スイッチを押していることを検出するときには、次順位に大きい類似度をもつ1つ又は複数の単語情報を認識結果として表示していくことで認識処理を実行するよう構成していた。なお、終端候補が取り消されることになる場合には、求められている類似度はすべてキャンセルされるよう処理されることになる。

第9図に、この従来の音声認識処理装置が実行するタイムチャートを図示する。ここで、図中の①は状態“S1”から状態“S2”に遷移したときに発生される音声区間の終端候補の検出信号を表し、②は状態“S2”から状態“S1”に遷移したときに発生される終端候補の取消信号を表し、③は状態“S0”から状態“S1”に遷移したときに検出される始端と、状態“S2”から状態

終端候補が終端として確定するかどうかを監視しており、閾値時間 L_{th} の間継続するときには音声区間が終了したと判断して状態“S0”に遷移する。このときの遷移時刻 t_3 が音声区間の終端の時刻として決定される。第8図に、音声パワーの時間経過に従って変化するこの状態遷移の一例を図示する。

従来の音声認識処理装置では、音声区間の終端候補が検出されることで仮の音声区間が検出されると、その仮の音声区間の特徴時系列データと単語辞書に管理される特徴時系列データとの間の類似性を表現する値の算出処理（以下、類似度に従う例で説明する）に入るとともに、この算出処理と並行して認識対象の単語を算出される類似度の大きさの順に従って並び変えるソーティング処理を実行していくことで、認識対象の単語との間の類似度を評価する処理を実行する。そして、この評価処理の終了後に検出されることになる音声区間の終端の検出時点において、上位の大きい類似度をもつ1つ又は複数の単語情報を認識結果とし

“S0”に遷移したときに検出される終端確定とにより区切られる音声区間の検出信号を表し、④は終端候補の検出に同期して実行される類似度の算出処理の実行状態を表す照合信号を表し、⑤はこの算出処理に同期して実行される単語のソーティング処理の実行状態を表すソーティング信号を表し、⑥は認識結果の単語情報の表示処理の実行状態を表す表示信号を表し、⑦は候補選択スイッチのオン・オフ状態を表すスイッチ信号を表している。

〔発明が解決しようとする課題〕

しかしながら、このような従来技術では、認識対象となる単語数、すなわち単語辞書に管理される単語数が例えば1千単語程度のオーダーの場合には、音声区間の終端候補の検出からその終端候補の終端としての確定までに要する時間である0.3秒程度の時間内に、すべての単語についての類似度を計算できることになることから何ら問題は起こらないものの、認識対象の単語数が数万単

語程度のオーダーとなると、すべての単語についての類似度を計算できるまでに数秒〜数十秒程度の時間がかかることになり、これがために、音声区間の終端時点では認識結果を表示することができないという問題点があった。すなわち、従来技術では、認識対象の単語数が超大語彙になると、第10図に示すように、音声区間の終端が検出されてから数秒〜数十秒程度も経過してからようやく認識結果の単語情報が表示されることになるという問題点があった。この問題点は認識対象の単語数が増えれば増える程大きな問題となることになる。ここで、図中の①ないし④は、第9図に示したものと同一信号を表している。

このように、従来技術では、認識対象の単語数が超大語彙になると、実用的な実時間の音声認識処理装置を構成できないという問題点があったのである。これを解決するために、ハードウェアを並列構成すること等で類似度や距離の計算時間を短縮していく方法を探ることも考えられるが、この方法を探ると膨大なハードウェアを用意しなく

置であって、認識された単語を表示するもの、5は検索要求手段であって、音声認識処理装置本体1に対して表示装置4上に次順位の認識結果の単語を表示するよう要求するもの、10は音響処理部であって、入力される音声信号の特徴時系列データを算出するとともに、該音声信号のパワー値を算出するもの、11は音声区間検出部であって、音声信号のもつパワー値に従って音声区間を検出するもの、12は照合部であって、検出された音声区間により抽出される音声信号の特徴時系列データと単語辞書2の特徴時系列データとを照合することで、入力される音声信号と認識対象の単語との間の類似性を評価するもの、13はソーティング部であって、評価された類似性に従い類似性の高い順に従って認識対象の単語の単語番号をソーティングするもの、14は一時メモリであって、ソーティング結果をメモリするもの、15は認識単語出力部であって、認識結果の1つ又は複数の単語を表示装置4に出力するもの、16は認識単語出力部15を構成する出力単語検索部であって、

てはならないという別の問題点をもたらすことになる。

本発明はかかる事情に鑑みてなされたものであって、マンマシン・インタフェース機能の改善を図ることで、非常に多くの単語情報をハードウェアの増加を招くことなく、かつオペレータに違和感を与えることなく実時間で認識できるようにする新たな音声認識処理装置の提供を目的とするものである。

(課題を解決するための手段)

第1図は本発明の原理構成図である。

図中、1は本発明を具備する音声認識処理装置本体であって、発声された音声信号に対応する単語を認識して出力するもの、2は単語辞書であって、認識対象の単語とその単語のもつ特徴時系列データとを単語番号を付して好ましくは使用頻度の大きい順に管理するもの、3はマイクロフォンであって、音声認識処理装置本体1に電気信号に変換された音声信号を入力するもの、4は表示装

置であって、一時メモリ14のメモリデータを参照することで出力すべき単語を検索するもの、17は認識単語出力部15を構成する検索起動部であって、出力単語検索部16に対して検索処理の起動を要求するもの、18は認識単語出力部15を構成する出力単語管理部であって、既に認識結果として出力された同一の音声信号に関しての単語情報を管理するもの、19は認識単語出力部15を構成する単語出力部であって、出力単語検索部16により検索された出力対象の単語を表示装置4に出力するものである。

(作用)

本発明では、オペレータがマイクロフォン3に対して発声することで音声信号が入力されると、音響処理部10は、入力されてきた音声信号の特徴時系列データを算出して照合部12に通知するとともに、その音声信号のパワーを検出して音声区間検出部11に通知するよう処理する。この通知を受けて、音声区間検出部11が音声信号のパ

ワーが判定レベル値以上に入ることによって音声区間の始端を検出した後に判定レベル値以下に下がることによって音声区間の終端候補を検出すると、照合部 12 は、入力されてくる特徴時系列データからその始端と終端候補との間の音声区間により区切られる特徴時系列データを抽出するとともに、単語辞書 2 の管理する特徴時系列データをその単語番号とともに読み出すよう処理する。

続いて、照合部 12 は、抽出した特徴時系列データと読み出した単語辞書 2 の特徴時系列データとを照合することで、単語辞書 2 の各単語がもつ音声信号との間の類似性を評価する処理を実行する。そして、ソーティング部 13 は、この照合部 12 の照合処理を受けて、評価された類似性の高い順に従って読み出した単語番号をソーティングして一時メモリ 14 に格納していく処理を実行する。この処理時にあって、音声区間検出部 11 が例えば 0.3 秒経過する前に音声信号のパワーが判定レベル値以上に転ずることによって検出した終端候補が本来の音声区間の終端でないことを検出すると、

単語番号を出力単語管理部 18 に登録する処理を実行する。

この認識結果の表示に対して、発声した単語が表示装置 4 上に表示されないときには、オペレータは、検索要求手段 5 に従って次の認識結果の単語を表示するよう要求する。この検索要求手段 5 からの検索指示を受け取ると、検索起動部 17 は、音声区間の終端の検出時と同様に出力単語検索部 16 に対して一時メモリ 14 の検索処理の起動を要求する。そして、この起動要求を受け取ると、出力単語検索部 16 は、一時メモリ 14 を参照することで、その時点までに求められている上位の高い類似性をもつ 1 つ又は複数の単語の単語番号を特定するとともに、この単語番号の内の出力単語管理部 18 に管理されていない単語番号を特定することで既出力の単語を除いたものを出力すべき単語として特定し、単語出力部 19 は、この特定された単語番号の単語を単語辞書 2 から読み出して表示装置 4 に出力し、続いて、出力単語検索部 16 は、この出力された単語番号を出力単語管

照合部 12 は類似性を評価するための照合処理を中止するとともに、ソーティング部 13 はソーティング処理を中止してそれまでに実行されたソーティング結果をクリアするよう処理する。

音声区間検出部 11 が終端候補から所定時間の間音声信号のパワー値が判定レベル値以下にあることによって音声区間の終端を検出すると、検索起動部 17 は、出力単語検索部 16 に対して一時メモリ 14 の検索処理の起動を要求する。ここで、この音声区間の終端が検出されるときにあって、照合部 12 は類似性の評価処理を実行し続けるとともに、ソーティング部 13 は単語番号のソーティング処理を実行し続けるよう処理する。検索起動部 17 から起動要求を受け取ると、出力単語検索部 16 は、一時メモリ 14 を参照することでその時点までに求められている上位の高い類似性をもつ 1 つ又は複数の単語の単語番号を特定し、単語出力部 19 は、この特定された単語番号の単語を単語辞書 2 から読み出して表示装置 4 に出力する。そして、出力単語検索部 16 は、この出力された

理部 18 に登録する処理を実行する。

そして、この表示にあって発声した単語が表示装置 4 上に表示されないときには、オペレータは、更に検索要求手段 5 に従って、続行されている類似性の評価処理を考慮した次の認識結果の単語を表示するよう要求していくことで、発声した単語を表示装置 4 上に表示させるよう処理していくことになる。

このように、本発明では、認識対象の単語数が多くなることで音声区間の終端の検出時点ですべての単語についての類似性が評価されないようなことがあっても、音声区間の終端の検出された時点で、それまでの照合処理により求められた類似性に従って認識結果を表示装置 4 に出力していくよう処理することから、発声の数秒〜数十秒経過後に認識結果が出力されるというような従来技術の欠点は解消されることになる。しかも、このとき、使用頻度の大きい順に従って単語の類似性を評価するよう構成することで、ほとんどの場合にあっては、この音声区間の終端の検出時点で発声

された音声信号に対しての認識処理が実現されることになる。そして、音声区間の終端の検出時点では評価されないような低頻度の単語については、オペレータが検索要求手段5を操作するまでの間に類似性が評価されることになるとともに、この新たに評価される類似性を含めたものでもって認識すべき単語が特定されて出力されるよう構成されることから、低頻度の単語にあっても、オペレータの違和感を招くことなく認識処理が実現されることになる。

更に、本発明では、音声区間検出部11により次の音声信号の音声区間の始端が検出されると、照合部12は実行中の照合処理を中止するとともに、ソーティング部13は実行中のソーティング処理を中止して一時メモリ14に格納されているソーティング結果をクリアする処理を実行する。この処理により、オペレータが発声した単語が表示装置4に表示されたことを確認して次の単語を発声してくることがあっても、直ちに、その入力されてくる単語の認識処理に入れるよう構成でき

第3図に示すように、認識対象の単語の単語読みラベル、特徴時系列データ、単語番号及び使用頻度を例えばアイウエオ順等の昇順に従って管理するもの、7は単語辞書ソーティング部であって、頻度付単語辞書6の管理データを使用頻度の大きい順にソーティングして、そのソーティングの順番に従って単語辞書2に対して認識対象の単語の単語読みラベル、特徴時系列データ及び単語番号を登録するものである。例えば、頻度付単語辞書6が第3図に示す管理データを管理するときには、単語辞書ソーティング部7は、第4図に示す単語の順番に従って認識対象の単語の単語読みラベル、特徴時系列データ及び単語番号を単語辞書2に登録する処理を実行する。これから、単語辞書ソーティング部7が動作すると、単語辞書2は、第5図に示すように、使用頻度の大きい順にソーティングされた認識対象の単語の単語読みラベル、特徴時系列データ及び単語番号を管理することになる。

音響処理部10は、第1図でも説明したように、

ることになる。ここで、次の音声信号の入力を指示するスイッチ手段を備える場合には、照合部12及びソーティング部13は、このスイッチ手段からの入力要求に応動して実行中の処理の中止を実行するという構成を採ることも可能である。

このように、本発明によれば、マンマシン・インタフェース機能の改善を図ることで、特別なハードウェアの追加をすることなく、非常に多くの単語情報をオペレータに違和感を与えることなく実時間で認識できるよう構成できることになる。

〔実施例〕

以下、実施例に従って本発明を詳細に説明する。

第2図に、本発明の一実施例を図示する。図中、第1図で説明したものと同じものについては同一の記号で示してある。5aは候補選択スイッチであって、第1図の検索要求手段5に相当するもの、20はパラメータバッファであって、音響処理部10により算出される特徴時系列データを一時的に記憶するもの、6は頻度付単語辞書であって、

入力される音声信号の特徴時系列データを算出するとともに、該音声信号のパワー値を算出するよう処理するものである。より具体的に説明するならば、数秒から数十秒の間隔に従って音声信号のパワー値の算出処理とスペクトル分析処理を実行することで、音声信号のパワー値とスペクトルの特徴時系列データとを算出するものである。スペクトル分析は、従来から用いられているアナログフィルタバンクによる方式、ディジタルフィルタバンクによる方式、高速フーリエ変換による方式等を適用することで実行される。このようにして算出されたパワー値は音声区間検出部11に出力されることになるとともに、特徴時系列データはパラメータバッファ20を介して照合部12に出力されることになる。

音声区間検出部11は、従来の音声認識処理装置の備える音声区間検出部と同様に、第7図に説明した状態遷移処理を実行して音声区間の始端、終端候補、終端候補取消及び終端を検出するもので、音声区間の始端を検出するときには照合部1

2に対して始端の検出信号を通知し、音声区間の終端候補を検出するときには照合部12に対して終端候補の検出信号を通知し、音声区間の終端候補取消を検出するときには、照合部12及びソーティング部13に対して終端候補の取消信号を通知し、音声区間の終端を検出するときには、認識単語出力部15に対して終端の検出信号を通知するよう処理することになる。この音声区間検出部11には、これらの信号の生成のために、第7図の状態遷移図で説明したところの閾値 P_{11} 及び L_{11} が外部から設定されるよう構成される。

そして、照合部12は、第1図でも説明したように、音声区間検出部11で検出される音声区間により抽出される音声信号の特徴時系列データと単語辞書2の管理する特徴時系列データとから距離や類似度等を算出することで、入力される音声信号と認識対象の単語との間の類似性を評価するよう処理するものである。例えば動的計画法等に従って類似性を評価することになる。

次に、第6図のタイムチャートを参照しつつ、

を実行することで特徴時系列データを算出してパラメータバッファ20に格納していくとともに、その音声信号のパワーを計算して音声区間検出部11に送出する処理を実行する。

この音声信号のパワーを受け取ると、音声区間検出部11は、入力されてくる音声パワーが閾値 P_{11} を超えることを検出することで音声区間の始端を検出するよう処理する。そして、この音声区間の始端を検出すると、音声区間に入ったことを表すために、認識単語出力部15に対して送出する終端確定信号(第6図中の④)を $show$ にリセットするとともに、照合部12に始端の検出信号パルスを送出する処理を実行する。続いて、音声区間検出部11は、音声パワーが閾値 P_{11} を下回することを検出することで音声区間の終端候補を検出するよう処理する。そして、この音声区間の終端候補を検出すると、仮の音声区間が検出されたことを表すために、照合部12に終端候補の検出信号パルス(第6図中の①)を送出する処理を実行する。

このように構成される本発明の実施例の動作処理について詳細に説明する。ここで、以下の説明にあつては、説明の便宜上、音声信号の特徴時系列データと単語辞書2の管理する特徴時系列データとの間の類似性の評価処理は類似度に従って実行することを想定している。また、第9図と同様に、第6図中の①は音声区間の終端候補の検出信号を表し、②は終端候補の取消信号を表し、③は始端と終端確定とにより区切られる音声区間の検出信号(認識単語出力部15に通知される終端確定信号となる)を表し、④は類似度の算出処理の実行状態を表す照合信号を表し、⑤はソーティング処理の実行状態を表すソーティング信号を表し、⑥は認識された単語情報の表示処理の実行状態を表す表示信号を表し、⑦は候補選択スイッチ5aのオン・オフ状態を表すスイッチ信号を表している。

オペレータがマイクロフォン3に対して発声することで音声信号が入力されると、音響処理部10は、入力されてきた音声信号のスペクトル分析

このようにして、仮の音声区間が検出されると、照合部12は、パラメータバッファ20からその検出された仮の音声区間の特徴時系列データを読み出し、更に、単語辞書2に管理されている特徴時系列データをその格納順(すなわち使用頻度の大きい単語順)に従って単語番号とともに読み出していくとともに、第6図の④に示すように、仮の音声区間の特徴時系列データと読み出していく単語辞書2の特徴時系列データとの間の類似度を算出していくよう処理する。そして、ソーティング部13は、この算出されていく類似度に従って、第6図の⑤に示すように、類似度の大きい順に類似度の求められた単語の単語番号をソーティングして一時メモリ14に格納していくよう処理する。この処理時にあって、音声区間検出部11が閾値 L_{11} の時間経過する前に音声信号のパワーが閾値 P_{11} を超えることで検出した終端候補が本来の音声区間の終端でないことを検出すると、第6図の②に示すように、照合部12及びソーティング部13に終端候補の取消信号パルス(第6図中の②)

を送出する処理を実行することで、照合部12の類似度の算出処理を中止させるとともに、ソーティング部13のソーティング処理を中止させてそれまでに実行されたソーティング結果をクリアさせるよう処理する。

音声区間検出部11が終端候補から閾値 P_{th} の時間の間音声信号のパワーが閾値 P_{th} 以下にあることで音声区間の終端を検出すると、音声区間検出部11は、認識単語出力部15に送出する終端確定信号(第6図中の③)をHighにセットする。この終端確定信号の立ち上がりにより、認識単語出力部15は音声区間の終端の検出通知を受け取ることになる。ここで、この音声区間の終端が検出されるときにあっても、単語辞書2に格納される単語数が多いので類似度の算出処理は継続して実行されている。このようにして、音声区間の終端の検出信号を受け取ると、認識単語出力部15は、一時メモリ14を参照することでその時点までに求められている例えば1位から20位までの大きな類似度をもつ単語番号を特定するととも

ように、この特定された単語番号の単語の単語読みラベルを単語辞書2から読み出して表示装置4に表示するよう処理する。

そして、この表示にあっても発声した単語が表示されないときには、オペレータは更に候補選択スイッチ5aをオンし、認識単語出力部15は、このオン処理に対応して同様の処理を繰り返していくことで、オペレータの発声した単語の単語読みラベルを表示装置4上に表示させるよう処理していくことになる。

このように、本発明では、使用頻度の大きい順に従って認識対象の単語のもつ類似度を算出していくとともに、音声区間の検出時点と要求のある時点において、その時点までに求められた類似度に従って認識結果を出力していくよう処理することから、高頻度の単語にあっては音声区間の検出時点で認識されるようになるとともに、低頻度の単語の単語にあってはオペレータの違和感を招くことなく認識処理が実現されるようになる。

このように、本発明では、超大語量の単語を認

に、第6図の⑥に示すように、この特定された単語番号の単語の単語読みラベルを単語辞書2から読み出して表示装置4に表示するよう処理する。このようにして表示される単語読みラベルは、使用頻度の大きい単語の中から認識されることから、ほとんどの場合、この音声区間の検出時点での認識処理に従って発声された音声信号の単語読みラベルが表示装置4に表示されることになる。

この認識結果の表示に対して、発声した単語の単語読みラベルが表示されないときには、オペレータは、第6図の⑦に示すように候補選択スイッチ5aをオンする。この候補選択スイッチ5aのオン操作は表示内容の確認を伴うことから通常数秒要することになる。このようにして、音声区間の終端の検出時点から数秒経過して候補選択スイッチ5aがオンされると、認識単語出力部15は、一時メモリ14を参照することでその時点までに求められている上位の大きな類似度をもつ単語番号であって既に出力されたものを除く単語番号を例えば20個特定するとともに、第6図の⑥に示す

認識対象として扱うことから、音声区間の終端確定時点においても類似度の算出処理を続行していく構成が採られることになる。一方、上述したように、本発明によればほとんどの場合音声区間の終端確定時点で認識が終了することになる。これから、本発明では、発声した単語読みラベルが表示されたことを確認することで、オペレータが続けて次の発声を行うことがあることを考慮して、音声区間検出部11が類似度の算出処理中に音声区間の始端を検出するときには、照合部12は実行中の類似度の算出処理を直ちに中止するとともに、ソーティング部13もまた実行中のソーティング処理を直ちに中止して一時メモリ14に格納されているソーティング結果をクリアする処理を実行するよう処理する構成を採る。これにより、次から次へと発声されてくる音声情報の認識処理を迅速に実行できるようになる。ここで、次の音声信号の入力を指示するスイッチ手段を備える場合には、このスイッチ手段のオン・オフモードを検出することで同様の処理に入るよう構成することを

可能である。

(発明の効果)

以上説明したように、本発明によれば、10万語以上というような超大語彙を実時間で認識できる音声認識処理装置を膨大なハードウェアを用いることなしに、かつ良好なマンマシン・インタフェースを確保しつつ実現できるようになる。そして、高頻度の単語については、従来通り音声区間の検出時点で認識されるようになるとともに、低頻度の単語についても認識対象とすることができることから、認識率の高い実用的な音声認識処理装置を提供できるようになるのである。

4. 図面の簡単な説明

第1図は本発明の原理構成図、

第2図は本発明の一実施例、

第3図は頻度付単語辞書の管理データの説明図、

第4図は単語辞書ソーティング部の処理内容の説明図、

第5図は単語辞書の管理データの説明図、

第6図は本発明の処理を説明する説明図、

第7図は音声認識処理装置が実行する音声区間検出処理機能の状態遷移図、

第8図は音声区間検出処理機能の状態遷移の一例、

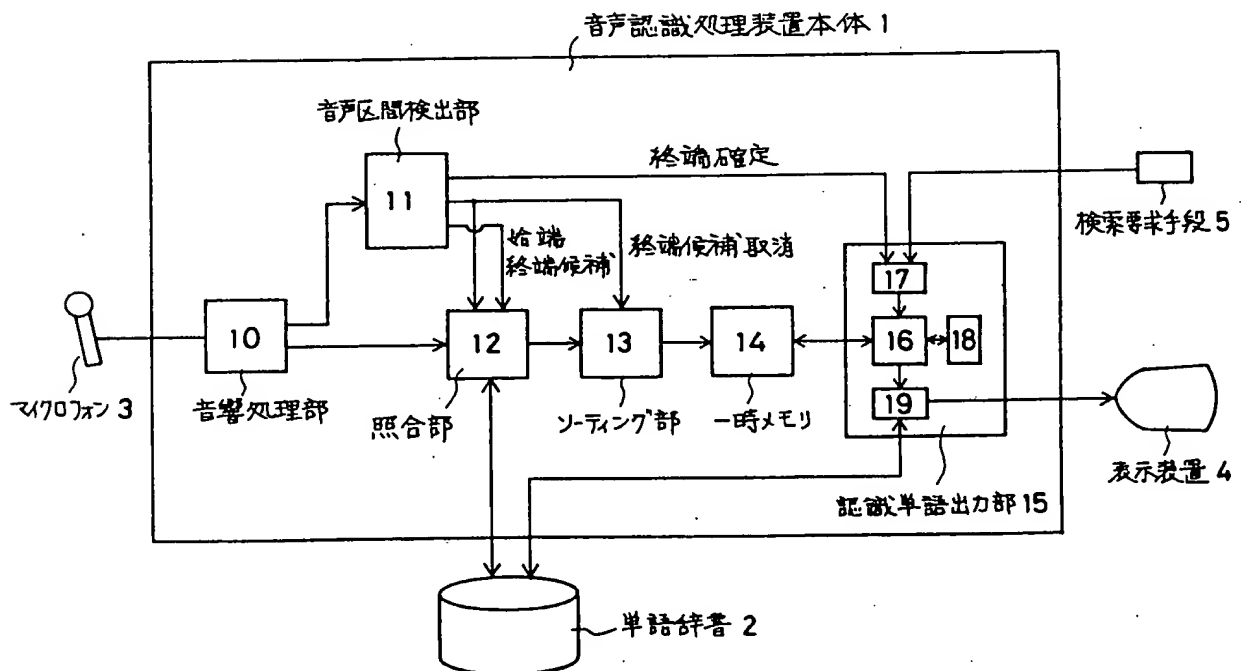
第9図は従来技術の説明図、

第10図は従来技術の問題点を説明する説明図である。

図中、1は音声認識処理装置本体、2は単語辞書、3はマイクロフォン、4は表示装置、5は検索要求手段、5aは候補選択スイッチ、6は頻度付単語辞書、7は単語辞書ソーティング部、10は音響処理部、11は音声区間検出部、12は照合部、13はソーティング部、14は一時メモリ、15は認識単語出力部、16は出力単語検索部、17は検索起動部、18は出力単語管理部、19は単語出力部、20はパラメータバッファである。

特許出願人 富士通株式会社

代理人 弁理士 森田 寛(外2名)

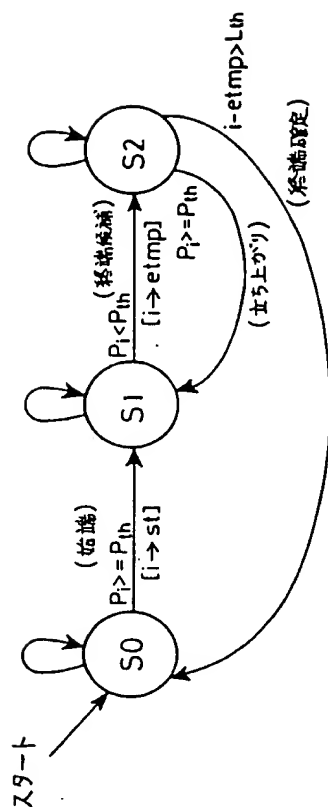


本発明の原理構成図

第1図

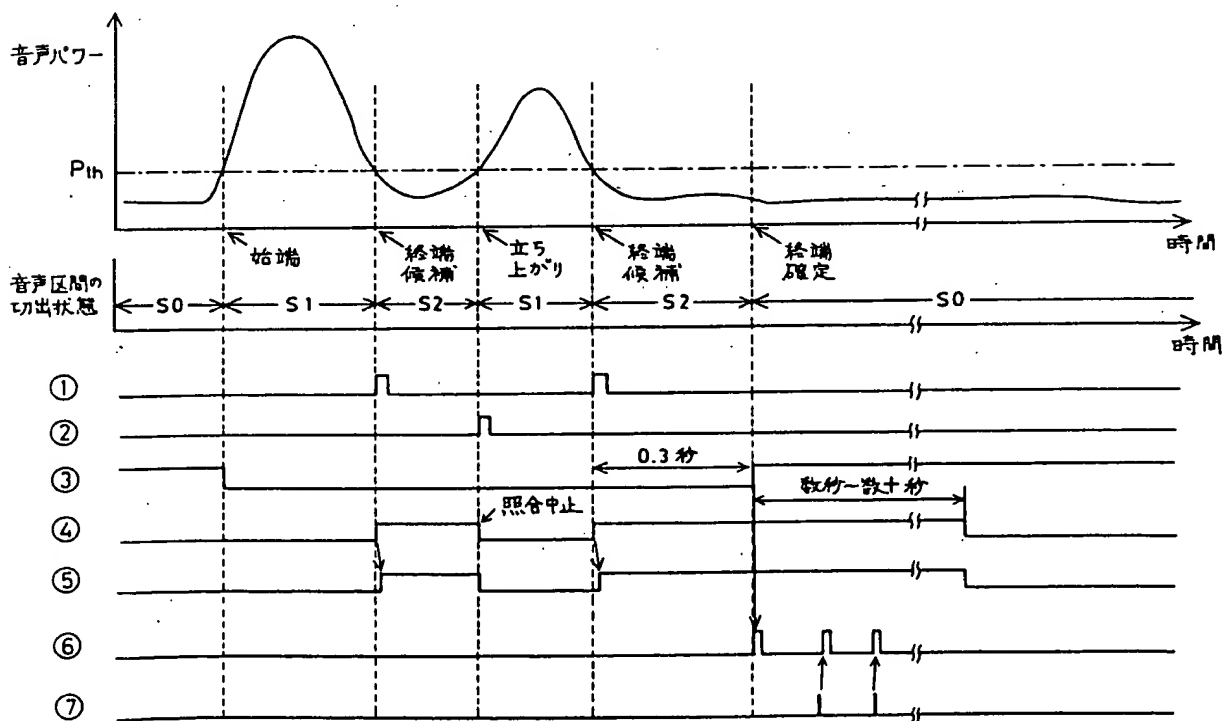
単語番号	単語読みラベル	特徴時系列データ
70	トウキョウ	
25	オオサカ	
1	アイチ	
2	アオモリ	
3	アキタ	

単語辞書の管理データの説明図
第 5 図



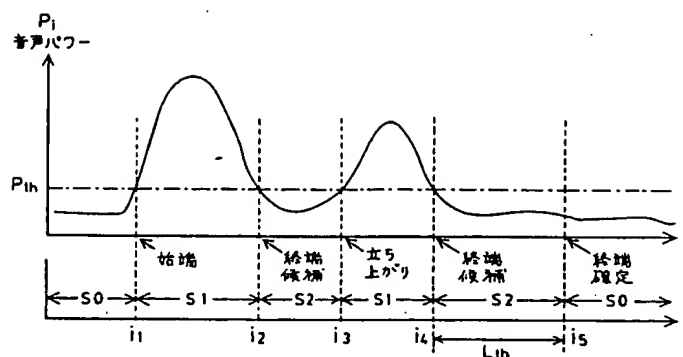
音声認識処理装置が実行する音声区間
検出処理機能の状態遷移図

第 7 図



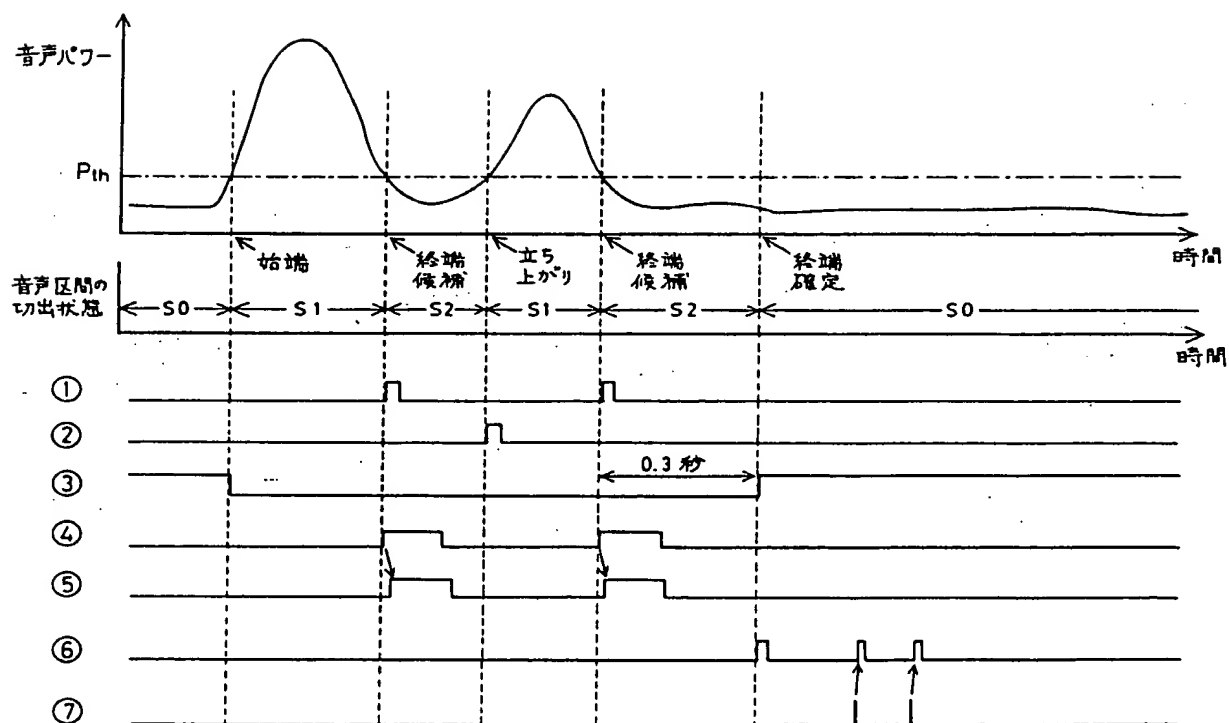
本発明の処理を説明する説明図

第 6 図



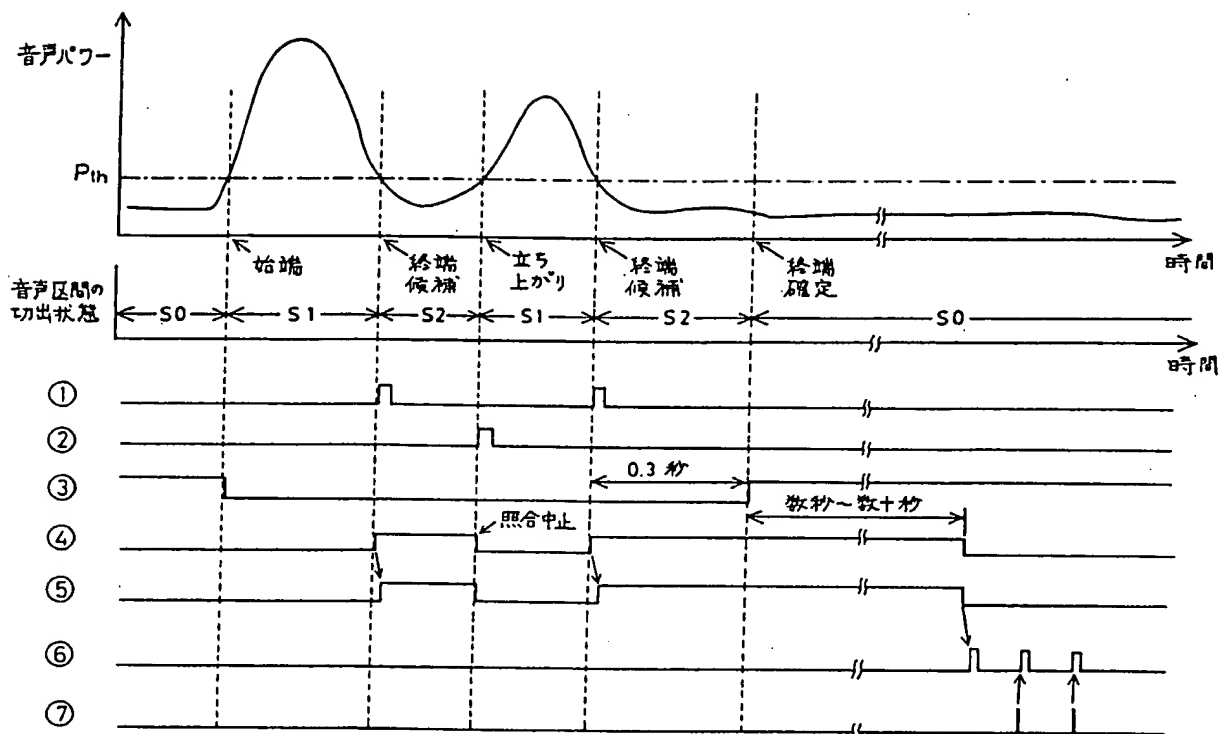
音声区間検出処理機能の状態遷移の一例

第 8 図



従来技術の説明図

第 9 図



従来技術の問題点を説明する説明図

第 10 図